Morphoproteomic characterization of lung squamous cell carcinoma fragmentation, a histological marker of increased tumor invasiveness

Casanova, Ruben ; Xia, Daniel ; Rulle, Undīne ; Nanni, Paolo ; Grossmann, Jonas ; Vrugt, Bart ; Wettstein, Reto ; Ballester, Rafael ; Astolfo, Alberto ; Weder, Walter ; Moch, Holger ; Stampanoni, Marco ; Beck, Andrew H ; Soltermann, Alex

Abstract: Accurate stratification of tumors is imperative for adequate cancer management. In addition to staging, morphological subtyping allows stratification of patients into additional prognostic groups. In this study, we used an image-based computational method on pan-cytokeratin immunohistochemical (IHC) stainings to quantify tumor fragmentation (TF), a measure of tumor invasiveness of lung squamous cell carcinoma (LSCC). In two independent clinical cohorts from tissue microarrays (TMA: n=208 patients) and whole sections (WS: n=99 patients), TF was associated with poor prognosis and increased risk of blood vessel infiltration. A third cohort from the cancer genome atlas (TCGA: n=335 patients) confirmed the poor prognostic value of TF using a similar human-based score on haematoxylin-eosin (HE) staining. Integration of RNA-seq data from TCGA and LC-MS/MS proteomics from WS revealed an upregulation of extracellular matrix remodeling and focal adhesion processes in tumors with high TF, supporting their increased invasive potential. This proposed histologic parameter is an independent and unfavorable prognostic marker that could be established as a new grading parameter for LSCC.

DOI: https://doi.org/10.1158/0008-5472.CAN-16-2363

Morphoproteomic characterization of lung squamous cell carcinoma fragmentation, a histological marker of increased tumor invasiveness

Authors

Ruben Casanova[1], Daniel Xia[2,3], Undine Rulle[1], Paolo Nanni[4], Jonas Grossmann[4], Bart Vrugt[1], Reto Wettstein[5], Rafael Ballester[5], Alberto Astolfo[6], Walter Weder[7], Holger Moch[1], Marco Stampanoni[6,8], Andrew H. Beck[9], Alex Soltermann[1]

1. Institute of Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland
2. Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
3. Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA
4. Functional Genomics Center Zurich, University/ETH Zurich, Zurich, Switzerland
5. Department of Informatics, University of Zurich, Zurich, Switzerland
6. TOMCAT Beamline, Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland
7. Division of Thoracic Surgery, University Hospital Zurich, Zurich, Switzerland
8. Institute for Biomedical Engineering, University/ETH Zurich, Zurich, Switzerland
9. Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School Boston, MA, USA

Running title

Morpho-proteomics of lung squamous cell carcinoma

Keywords

Lung, squamous cell carcinoma, digital pathology, prognosis, label-free mass spectrometry

Financial support

This work was supported by grants from the Swiss Cancer League (reference number F-87701-31-01) and the Swiss National Science Foundation SystemsX (reference number M-87704-01-02) to A. Soltermann.

Corresponding author

Ruben Casanova, Institute of Pathology and Molecular Pathology, University Hospital Zurich, Schmelzbergstrasse 12, CH-8091 Zurich, Switzerland. Tel: +41 44 255 29 21; ruben.casanova@usz.ch

Conflict of interest

The authors declare to have no competing financial interest

Abstract

Accurate stratification of tumors is imperative for adequate cancer management. In addition to staging, morphological subtyping allows stratification of patients into additional prognostic groups. In this study, we used an image-based computational method on pan-cytokeratin immunohistochemical (IHC) stainings to quantify tumor fragmentation (TF), a measure of tumor invasiveness of lung squamous cell carcinoma (LSCC). In two independent clinical cohorts from tissue microarrays (TMA: n=208 patients) and whole sections (WS: n=99 patients), TF was associated with poor prognosis and increased risk of blood vessel infiltration. A third cohort from the cancer genome atlas (TCGA: n=335 patients) confirmed the poor prognostic value of TF using a similar human-based score on haematoxylin-eosin (H&E) staining. Integration of RNA-seq data from TCGA and LC-MS/MS proteomics from WS revealed an upregulation of extracellular matrix remodeling and focal adhesion processes in tumors with high TF, supporting their increased invasive potential. This proposed histologic parameter is an independent and unfavorable prognostic marker that could be established as a new grading parameter for LSCC.

Introduction

Lung squamous cell carcinoma (LSCC) is the second most frequent histological subtype of non-small cell lung carcinoma (NSCLC), accounting for 25-30% of all lung cancers in Europe (1). According to the 2015 WHO classification, LSCCs are further separated into keratinizing and non-keratinizing subtypes. However, the prognostic relevance of this subtyping remains unclear (2,3). Stratification of lung cancers is an imperative step for adequate disease management and is primarily achieved by TNM staging (primary tumor extent, lymph nodes status and distant metastases). In addition, tumor grading based on morphological parameters such as overall architecture or cell and nuclear pleomorphism also allows stratifying patients into prognostic groups. However, there is still no well-established grading system for LSCC (4).

Tumor invasion is supported by *de novo* formation of desmoplastic stroma, which provides not only physical support to cancer cells but also favors tumor expansion and invasion as a net effect of tumor-stroma inter-talk (5). LSCC invasion is histologically characterized by tumor clusters of variable sizes surrounded by such specialized stroma. This is reflected by an apparent fragmentation of both central and peripheral portions of the tumor mass. The quantification of such tumor fragments on whole histological sections may be used as a metric for tumor aggressiveness which could serve as a novel grading parameter to stratify LSCC into prognostic groups.

In this study, we performed an image-based computational analysis for unbiased tumor fragmentation (TF) quantification and applied extended the method with a human-based scoring system relevant for pathologists. Additionally, we integrated tumor histology with molecular data to evaluate the biological processes associated to this morphological feature. A retrospective study on three independent clinical cohorts was performed to assess the potential prognostic relevance of TF as novel grading parameter for LSCC.

Materials and Methods

Patient cohorts

In this retrospective study, consecutive patients with surgical resection of their primary lung SCC at the University Hospital Zurich were selected. Squamous cell differentiation was reviewed on haematoxylin-eosin (H&E) and alcian blue-periodic acid schiff (AB-PAS) stains. Differential diagnosis of non-keratinizing SCC versus solid adeno-, null-phenotype or neuroendocrine large cell carcinoma was performed by respective IHC using TTF1, p40, p63, CK7 and synaptophysin antibodies. Cases with mixed or unclear histology were excluded. Patients with synchronic or metachronic second primary tumor, in particular head and neck SCC, with neo-adjuvant chemo-/radiotherapy or with overall survival (OS) < 1 month post-surgery were also excluded. The TMA cohort consisted of 233 lung SCC patients between January 1993 and December 2002, as described (6). Patients were further excluded if the quality of the pan-cytokeratin IHC was insufficient for automatic image analysis on both cores, e.g. due to presence of non-cancerous tissue, more than 80% stroma or major cutting artefacts. In total 208 patients met inclusion criteria. Adjuvant treatment was administered to 75 (36%) patients from which 62 (82%) received radiotherapy, 8 (11%) chemotherapy and 5 (7%) both. The WS cohort consisted of non-redundant 167 patients between January 2003 and February 2010 having complete clinical data. After revision for exclusion criteria, 99 patients were included. Adjuvant therapy was administered to 30 (30%) of the patients from which 6 (20%) received radiotherapy, 20 (67%) chemotherapy and 4 (13%) both. TCGA clinical data was downloaded from cBioPortal for Cancer Genomics, Memorian Sloan Kettering Cancer Center (7) for 504 lung SCC patients. After exclusion due to OS<1 month, neo-adjuvant chemotherapy or insufficient image quality, 335 patients remained. Thirty-one (9%) patients had a documented adjuvant therapy whereby 3 (10%) received radiotherapy, 21 (68%) targeted therapy and 7 (22%) both. Our study was approved by the Ethical Commission of the Canton of Zurich under reference number KEK ZH-Nr. 29-2009/14.

Tissue microarray construction and immunohistochemistry

For the TMA cohort, two representative paraffin blocks with tumor were selected. One tissue core of 600µm diameter was taken from each tissue block and transferred into a recipient block. For the whole sections cohort, two blocks representing best the overall tumor morphology were selected and 2µm thick sections were cut from each block. Immunohistochemistry (IHC) was performed on an automated platform (Ventana Medical Systems, Tucson, AZ, USA) using mouse monoclonal anti-human cytokeratin AE1/AE3 (M3515, DAKO, dilution 1:50), rabbit polyclonal Periostin (Abcam, 1:1000) and mouse monoclonal Versican (2B1, Seikagaku, 1:500). Detection was completed with a respective secondary antibody and the OptiView DAB kit (Ventana Medical Systems).

Image acquisition

Immunohistochemically stained sections were scanned with a high resolution whole slide scanner (Nanozoomer Digital Pathology, Hamamatsu, Japan) using a 40x objective with spatial resolution of 0.23µm/pixel. TMA and whole sections images were analyzed with a spatial resolution of 1.84µm/pixel and 9.2µm/pixel, respectively. Whole sections were further annotated by a surgical pathologist (A.S.) in order to select tumor tissue. The surrounding non-tumor lung tissue was excluded from the analysis. A formalin-fixed cylindrical tissue sample of 1 cm diameter from a pT3 lung SCC was imaged by X-ray microtomography at the beamline for tomographic microscopy and coherent radiology experiments (TOMCAT), Paul Scherrer Institute (PSI, Würenlingen, Switzerland). Sample was scanned using a microscope equipped with a PCO-Edge camera mounted with a 10x objective for a spatial resolution of 0.65µm /pixel using propagation-based X-ray phase contrast CT, as described (8).

Image processing

Automatic morphometric analysis was performed using FIJI (9). Color-based segmentation allowed separating the tumor tissue (brown) from its surrounding stroma (blue/grey). Color threshold values were previously validated by the pathologist. A similar approach for tumor

segmentation by cytokeratin stained sections has been described for colon carcinoma (10). The segmentation algorithm included mean shift and median filtering, color thresholding using the *L*a*b* color space and particle size filtering with a minimum size threshold of 100µm^2, considered as noise. The resulting segmentation masks were used for automated scoring of tumor fragmentation.

Tumor fragmentation scoring

Tumor fragments refer to tumor clusters completely separated from each other on a two-dimensional plane by intervening desmoplastic stroma. Tumor fragmentation was automatically scored on the pan-CK stainings of both TMA and WS by counting the total number of disconnected particles larger than 800µm^2 (circa ≥ 5 cells) for each segmentation mask. TF scores were summed up over two images when available or duplicated for cases with single images. In addition, 88/99 WS patients were also represented on a secondary TMA from which 77 had sufficient image quality for automated TF scoring. TF scores from this subset were used to address scoring heterogeneity between TMA cores and whole sections for the same tumor. Human-based TF scoring was done by one observer (obs1) on H&E stained tissues. On TMA cores, TF was scored using all available magnifications and on the area of highest fragmentation for the TCGA and WS cohorts, under a magnification of 50x. This corresponded to 1920x1036 pixels (3.5x1.9mm) for TCGA image frames and to 4mm FOV for WS with light microscopy. Two additional observers (obs2-3) evaluated 20% of the TCGA images (n=67) to address inter-observer variability.

Sample preparation for mass spectrometry

A total of 48 samples from the whole sections cohort were selected for mass spectrometry based on the availability of at least 1 tumor paraffin block containing >80% carcinoma epithelia per total tissue surface. Two 20µm thick microtome cuts were deparaffinized in xylene and washed with 96% ethanol. Samples were suspended in 120µl of SDS buffer (4%

SDS, 100mM Tris / HCL pH 8.2, 0.1M DTT – dithiothreitol) and boiled at 95°C for 20 minutes followed by 2h at 80°C and processed with High Intensity Focused Ultrasound (HIFU) for 10 min, setting the ultrasonic amplitude to 65%. Protein concentration was determined using the Qubit® Protein Assay Kit (Life Technologies, Zurich, Switzerland). For each sample, 20µg protein were taken for on-filter digestion using an adaptation of the filter-aided sample preparation (FASP) protocol (11). Briefly, proteins were diluted in 200µl of UT buffer (Urea 8M in 100mM Tris/HCL pH 8.2), loaded on Ultracel 30000 MWCO centrifugal unit (Amicon Ultra, Merck, Darmstadt, Germany) and centrifuged at 14000g. SDS buffer was exchanged by one centrifugation round of 200µl UT buffer. Alkylation of reduced proteins was carried out by 5 min incubation with 100µl iodoacetamide 0.05M in UT buffer, followed by three 100µl washing steps with UT and three 100µl washing steps with NaCl 0.5M. Finally, proteins were on-filter digested using 120µl of 0.05 triethylammonium bicarbonate buffer (pH 8) containing trypsin (Promega, Madison, WI, USA) in ratio 1:50 (w/w). Digestion was performed overnight in a wet chamber at room temperature. After elution, the solution containing peptides was acidified to a final 0.1% TFA, 3% acetonitirile concentration. Peptides were desalted using self-packed C18 Stage-Tips, dried and re-solubilized in 15µl of 3% acetonitrile, 0.1% formic acid for MS analysis.

Liquid chromatography-mass spectrometry analysis

Mass spectrometry analysis was performed on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled to an Eksigent-Nano-HPLC system (Sciex, Framingham, MA, USA). Solvent composition at the two channels was 0.1% formic acid for channel A and 0.1% formic acid, 99.9% acetonitrile for channel B. For each sample 2µL of peptides were loaded on a self-made column (75µm × 150mm) packed with reverse-phase C18 material (ReproSil-Pur 120 C18-AQ, 1.9µm, Dr. Maisch GmbH, Ammerbuch, Germany) and eluted at a flow rate of 300nL/min by a gradient from 3 to 25% B in 65 min, 35% B in 5 min and 97% B in 5 min. Samples were acquired in a randomized order. The mass

spectrometer (Tune page v1.1) was configured to fragment peptide precursor ions in data-dependent mode, allowing a maximum of 3 s between the full-scan spectra (top speed mode). Full-scan MS spectra (300–1500 m/z) were acquired at a resolution of 120000 at 200 m/z after accumulation to an automated gain control (AGC) target value of 400000. Wide quadrupole isolation was used, and an injection time of 50ms was set. Precursors with an intensity above 5000 were selected for MS/MS. Ions were isolated using a quadrupole mass filter with 1.6 m/z isolation window and fragmented by higher-energy collisional dissociation (HCD) using a normalized collision energy of 30. Fragments were detected in the linear ion trap using adapted "Universal Method" settings: the scan rate was set to Rapid, the automatic gain control was set to 100 ions and the maximum injection time was 250 milliseconds. Charge state screening was enabled and singly and unassigned charge states were rejected. Precursor masses previously selected for MS/MS measurement were excluded from further selection for 25 seconds, and the exclusion window was set at 10 ppm. The samples were acquired using internal lock mass calibration on m/z 371.1010 and 445.1200.

Protein identification and label free protein quantification

Protein label-free quantification was performed using the software Progenesis QI for proteomics (v.4.0.4265.42984) software (Nonlinear Dynamics, Newcastle upon Tyne, UK), using as reference for alignment the raw-file of the pool sample. Normalization was kept with default settings. From each Progenesis peptide ion (default sensitivity in peak picking) a maximum of the top five tandem mass spectra were exported into a Mascot generic file using charge deconvolution and deisotoping option and a maximum number of 200 peaks per MS/MS. This Mascot generic file (mgf) was searched with Mascot 2.4.3.3 (Matrix Science Ltd., London, UK) against the forward Uniprot database for *Homo sapiens*, concatenated to a reversed decoyed FASTA database and 260 common mass spectrometry protein contaminants. The parameters for precursor tolerance and fragment ion tolerance were set to ± 10 ppm and ± 0.6 Da. Enzyme was set to trypsin and one missed cleavage was allowed.

Carbamidomethylation of cysteine was set as fixed modification, while oxidation (M) and deamidation (N, Q) were set as variable. The resulting .dat file was loaded into Scaffold v4.1.1 (Proteome Software) and filtered at peptide and protein False Discovery Rates (FDRs) were set to 1% and 9%, respectively. Finally, the Scaffold Spectrum Report was imported back into Progenesis. For quantification, all proteins identified with at least two peptide ions were assessed, resulting in an estimated protein FDR of 0.5%. Proteins were grouped with Progenesis and the relative quantification using Hi-N (N=3) peptides was used. For protein quantification, the average of the normalized abundance from the most intense N peptide ions of each protein group was calculated individually for each sample. This generated the normalized quantitative protein abundance. Protein levels were further log2 transformed for statistical testing. Differentially expressed proteins were identified by Significance Analysis of Microarrays as described (12). TF scores were log2 transformed and addressed as quantitative response using the standard regression method on median centered protein levels. Proteins significantly correlated with TF scores (FDR<0.05) were considered for further analysis. Gene ontology enrichment analysis was performed using WEB-based GEne SeT AnaLysis Toolkit (13) using default settings. All identified proteins (n=2614) were used as background (Table S1).

Gene ontology enrichment networks and pathway analysis for TCGA mRNAseq data

Analysis-ready standardized TCGA mRNA-Seq data was downloaded from the Broad GDAC Firehose stddata__2015_11_01 run. Two groups showing clearly distinguishable TF profiles on TCGA H&E stainings, consisting of patients with the 25% lowest (TF <6) and 25% highest (TF >18) scores were selected for mRNA differential expression analysis. For statistical testing, the R-package EdgeR was used to compare them. In total mRNAseq data for 20531 genes was available. Genes with a minimum of 1 count-per-million, in at least half of the samples, were selected (n=14151). Default parameter settings were applied. Candidates having a FDR <0.05 were included in the gene ontology enrichment analysis. Gene ontology

enrichment map was generated with the software Cytoscape (14) using BINGO (15) and Enrichment Map (16) plugins. Genes upregulated in high and low TF tumors, respectively, were uploaded separately in BINGO. Enrichment analysis was performed with the default hypergeometric test using a significance level <0.05 after Bonferroni correction for multiple testing. Enrichment map was generated using BINGO outputs for each tumor morphotype with the following parameter: p-value cutoff <0.001, FDR <0.05, Jaccard coefficient=0.25. The generated network is represented by nodes (gene sets), edges (mutual overlap) and colors (tumor morphology groups: blue=low TF, red=high TF). Pathway analysis was done using the KEGG database (17) through WEB-based GEne SeT AnaLysis Toolkit.

Data interpretation and statistical analysis

All statistical analyses were performed on SPSS version 22 software (SPSS Inc., Chicago, USA). Relapse-free survival was measured from the date of surgery to the date of documented relapse or death as described (18). RFS was assessed only for patients who presented no evidence of remaining tumor (incomplete resection and/or metastases) after surgery. Kaplan Meier survival curves were evaluated using log-rank tests using TF scores dichotomized at the median. Hazard ratios were assessed by Cox regression. Clinical correlations were addressed using the non-parametric spearman's rank correlation test. The association of clinical parameters with survival was computed by univariate Cox regression. In addition, clinically relevant parameters were introduced into multivariate Cox regressions. P-values <0.05 were considered significant.

Results

LSCC shows a variable degree of tumor fragmentation

Tumor fragmentation (TF) scores were automatically computed as the total number of tumor fragments in pan-cytokeratin stained TMAs of 208 LSCC patients and whole sections (WS) of 99 patients (Table 1). The size of tumor fragments within individual tumors varied considerably (Figure 1A). By using X-ray microtomography we saw that tumor fragments

represent projections of interconnected epithelial branches from a three dimensional perspective (Figure 1B). The distribution of the computed TF scores was comparable in the TMA and WS cohorts (Figure S1A-B left). The assessment of a subset of 77 patients from WS with matching TMA cores from a secondary TMA, showed a positive correlation of TF scores between whole sections and tumor cores from the same tumor (c.coeff=0.484, $p<0.001$). However, due to intra-tumor heterogeneity the WS analysis allows a more reliable assessment of TF scores as demonstrated by the correlation analysis of computed TF scores from two tumor regions (Figure S1A-B right). We further tested the applicability of TF scoring by human eye using H&E stained images of the TMA and WS cohorts, as well as in an external TCGA LSCC cohort with similar clinical characteristics. The distribution of TF scores was comparable between the computed and human-based evaluations (Figure S1C-E). Histological examples of tumors with high and low fragmentation from TMA, WS and TCGA cohorts are shown in Figure 2.

Tumor fragmentation is associated with increased invasiveness and worse outcome

Clinico-pathologic correlations showed that high TF was consistently associated with vessel infiltration on both TMA and WS (Table S2). Perineural and mediastinal invasion were further evaluated on WS and positively correlated with TF as well ($p=0.001$, $p=0.004$ respectively). Survival analysis showed that high TF is a poor prognostic factor for OS and was confirmed using the external TCGA cohort with human-based scores (Figure 3). Increasing TF scores had a significantly poor impact on OS as showed by both univariate and multivariate analysis (Table 2). Analysis for relapse free survival showed comparable results for the three patient cohorts (Table S3). Human-based scores showed similar results for WS (Table S4). By evaluating the subset of 77 patients from WS with matching TMA cores, we observed a decrease of prognostic power using TMA cores (OS/RFS: $p=0.060/0.152$), in comparison with whole sections (OS/RFS: $p<0.001/<0.001$), possibly due to the higher impact of intra-tumor heterogeneity on the TMA. TF scores using different minimal fragment

size thresholds were also evaluated on WS. The prognostic relevance of TF was similar for a wide range of size cutoffs (5-100 cells) for both OS and RFS (Figure S2).

Tumor fragmentation is associated with changes in extracellular matrix

We selected TCGA patients with the upper and lower quartiles of TF scores to identify molecular characteristics associated with TF morphotypes and identified 910 genes associated with the high TF. Enriched biological processes involved tissue development, extra-cellular matrix (ECM) organization and cell adhesion processes (Figure S3). KEGG analysis notably showed an upregulation of pathways involved in ECM-receptor interaction, focal adhesion and protein digestion. In parallel, mass spectrometry-based label-free quantification identified 154 proteins significantly associated with increased TF (Table S5) including extracellular (20%) and cytoskeletal (15%) components. In accordance with mRNA-Seq analysis, we observed an enrichment of genes involved in developmental processes.

In contrast, low TF was associated with 554 genes involved in diverse metabolic pathways including xenobiotics metabolism (Figure S3), and 205 proteins representing in majority intracellular (70%) and nuclear proteins (46%). Biological nodes associated with decreasing TF were mostly related to hormone regulation processes.

We validated the expression of Periostin and Versican, two ECM proteins associated with TF in both RNA and protein analyses, by immunohistochemistry (IHC). The expression of both proteins in the tumor and stromal compartments was significantly associated with TF (Figure 4).

Discussion

LSCC is histologically characterized by a variable degree of keratinization and/or the formation of intercellular bridges between tumor epithelial cells. Collective cell migration is predominant in LSCC as other invasion modes such as EMT (epithelial-mesenchymal transition, fibroblastoid) or amoeboid are less observed probably due to the tight desmosomal

adhesions (19,20). This invasion pattern results in an apparent network of epithelial branches in 3D which is histologically depicted by a high variability of epithelia sizes in 2D. In this study, we used an image-based computational method to quantify such tumor fragments, hypothesizing that high fragmentation is a trait of increased tumor invasiveness. In support of this hypothesis, we showed that TF was associated with increased blood vessel, mediastinal and perineural invasion and worse patient outcome. Consistently, molecular analysis showed an upregulation of processes involved in ECM remodeling and focal adhesion, major characteristics of increased cellular motility (20-23). The association of TF with two identified ECM proteins, Periostin and Versican, was validated by immunohistochemistry. Periostin is a secreted ECM protein is observed in the desmoplastic stroma of a variety of cancers and known to promote cell invasion (24-27). It is also involved in cardiac remodeling after myocardial infarction (28). Versican is a major proteoglycan of the extracellular matrix upregulated in several tumor types including lung cancer (29) and has been notably associated with poor prognosis in NSCLC (30). It is also shown to favor tumor metastasis in Lewis lung carcinoma cell lines (31). In contrast, tumors with low TF showed an enrichment of diverse metabolic and xenobiotic processes notably contributing to the regulation of a number of chemotherapeutic drugs (32-34).

TNM staging is the only established system for predicting LSCC prognosis, whereas the value of tumor grading along the keratinization qualifier remains unclear. In this study, we evaluated the clinical relevance of LSCC stratification based on TF. Survival analysis showed that high TF is a poor prognostic marker independent from stage. This could be a useful additional grading parameter suggesting tumor invasiveness. Alternative histological parameters such as tumor budding have been proposed as various measures of tumor invasion in cancers notably in colorectal carcinoma (35-37). In NSCLC, single cells, tumor buds (<5 cells) or nests (≤15 cells) together with stroma thickness have also shown an impact on patient's survival (2,3,38,39), although small invasive tumor clusters are less frequent in

LSCC. As shown in Figure 1, the typical median and average LSCC epithelia clusters size in 2D is at least one order of magnitude higher than single cells and tumor buds. Furthermore, studies of the 3D tumor microarchitecture suggest that single cells may be part of tumor buds which in turn may belong to larger tumor nests or branches (40).

Alternatively, current progresses in digital pathology have led to automated identification of prognostic features on histological sections, notably for non-small cell lung cancer (41). However, such methods are mostly based on the quantification of cellular (mostly nuclear and cytoplasmic) features. Our computational method is focused on a higher order feature of the tumor epithelia, related to tumor invasion patterns which can be favorably translated into a human-based scoring system on H&E tissue staining. The proposed computational image-based analysis allowed unbiased scoring of tumor fragments, but has nevertheless intrinsic limitations. The main drawback is that normal epithelial lung structures as well as necrotic residues are also immune-reactive with pan-cytokeratin. To minimize false positive tumor fragments, adjacent normal tissue was covered and a minimal size threshold for tumor fragments (>800μm^2) was set. Finally, the clinical relevance of TF in the context of adjuvant therapy could not be addressed in this study, because of the high heterogeneity of treatment modalities across cohorts.

In conclusion, we have shown using an image-based computational approach, that high fragmentation of LSCC is a histologic trait associated with increased aggressiveness. In addition, the integration of molecular data showed an upregulation of proteins favoring extracellular matrix remodeling and focal adhesion, supporting the increased invasive potential of tumors with such high fragmentation. The proposed histologic parameter is an independent unfavorable prognostic marker that could be envisaged as new grading parameter for LSCC.

Acknowledgements

We would like to acknowledge Dr. Karina Silina (Institute of Experimental Immunology), Prof. Bernd Bodenmiller (Institute of Molecular Life Sciences), Prof. Renato Pajarola (Department of Informatics), Dr. Dimitri Korol (Cancer Registry), Dr. Peter Schraml (Institute of Pathology and Molecular Pathology) for their critical input.

References

1. Francisci S, Minicozzi P, Pierannunzio D, Ardanaz E, Eberle A, Grimsrud TK, et al. Survival patterns in lung and pleural cancer in Europe 1999-2007: Results from the EUROCARE-5 study. Eur J Cancer 2015; pii: S0959-8049(15)00713-3.
2. Kadota K, Nitadori J, Woo KM, Sima CS, Finley DJ, Rusch VW, et al. Comprehensive pathological analyses in lung squamous cell carcinoma: single cell invasion, nuclear diameter, and tumor budding are independent prognostic factors for worse outcomes. J Thorac Oncol 2014;9(8):1126-39.
3. Weichert W, Kossakowski C, Harms A, Schirmacher P, Muley T, Dienemann H, et al. Proposal of a prognostically relevant grading scheme for pulmonary squamous cell carcinoma. Eur Respir J 2016;47(3):938-46.
4. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JH, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol 2015;10(9):1243-60.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646-74.
6. Soltermann A, Tischler V, Arbogast S, Braun J, Probst-Hensch N, Weder W, et al. Prognostic significance of epithelial-mesenchymal and mesenchymal-epithelial transition protein expression in non-small cell lung cancer. Clin Cancer Res 2008;14(22):7430-7.
7. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2(5):401-4.

8. Astolfo A, Lathuiliere A, Laversenne V, Schneider B, Stampanoni M. Amyloid-beta plaque deposition measured using propagation-based X-ray phase contrast CT imaging. J Synchrotron Radiat 2016;23:813-9.

9. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9(7):676-82.

10. Hershkovitz T, Shenhav A, Sabo E, Ben-Izhak O, Hershkovitz D. Development of a computerized morphometry application for assessment of the tumor fraction in colon carcinoma tissue samples. Appl Immunohistochem Mol Morphol 2013;21(1):54-8.

11. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods 2009;6(5):359-62.

12. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98(9):5116-21.

13. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013;41:W77-83.

14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498-504.

15. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005;21(16):3448-9.

16. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;5(11):e13984.

17. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;44(D1):D457-62.

18. Punt CJ, Buyse M, Kohne CH, Hohenberger P, Labianca R, Schmoll HJ, et al. Endpoints in adjuvant treatment trials: a systematic review of the literature in colon cancer and proposed definitions for future trials. J Natl Cancer Inst 2007;99(13):998-1003.

19. Yilmaz M, Christofori G, Lehembre F. Distinct mechanisms of tumor invasion and metastasis. Trends Mol Med 2007;13(12):535-41.

20. Friedl P, Alexander S. Cancer invasion and the microenvironment: plasticity and reciprocity. Cell 2011;147(5):992-1009.

21. Yamaguchi H, Wyckoff J, Condeelis J. Cell migration in tumors. Curr Opin Cell Biol 2005;17(5):559-64.

22. Marx J. Cell biology. Podosomes and invadopodia help mobile cells step lively. Science 2006;312(5782):1868-9.

23. Friedl P, Wolf K. Tube travel: the role of proteases in individual and collective cancer cell invasion. Cancer Res 2008;68(18):7247-9.

24. Gillan L, Matei D, Fishman DA, Gerbin CS, Karlan BY, Chang DD. Periostin secreted by epithelial ovarian carcinoma is a ligand for alpha(V)beta(3) and alpha(V)beta(5) integrins and promotes cell motility. Cancer Res 2002;62(18):5358-64.

25. Yan W, Shao R. Transduction of a mesenchyme-specific gene periostin into 293T cells induces cell invasive activity through epithelial-mesenchymal transformation. J Biol Chem 2006;281(28):19700-8.

26. Butcher JT, Norris RA, Hoffman S, Mjaatvedt CH, Markwald RR. Periostin promotes atrioventricular mesenchyme matrix invasion and remodeling mediated by integrin signaling through Rho/PI 3-kinase. Dev Biol 2007;302(1):256-66.

27. Soltermann A, Ossola R, Kilgus-Hawelski S, von Eckardstein A, Suter T, Aebersold R, et al. N-glycoprotein profiling of lung adenocarcinoma pleural effusions by shotgun proteomics. Cancer 2008;114(2):124-33.

28. Shimazaki M, Nakamura K, Kii I, Kashima T, Amizuka N, Li M, et al. Periostin is essential for cardiac healing after acute myocardial infarction. J Exp Med 2008;205(2):295-303.

29. Isogai Z, Shinomura T, Yamakawa N, Takeuchi J, Tsuji T, Heinegard D, et al. 2B1 antigen characteristically expressed on extracellular matrices of human malignant tumors is a large chondroitin sulfate proteoglycan, PG-M/versican. Cancer Res 1996;56(17):3902-8.

30. Pirinen R, Leinonen T, Bohm J, Johansson R, Ropponen K, Kumpulainen E, et al. Versican in nonsmall cell lung cancer: relation to hyaluronan, clinicopathologic factors, and prognosis. Hum Pathol 2005;36(1):44-50.

31. Kim S, Takahashi H, Lin WW, Descargues P, Grivennikov S, Kim Y, et al. Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. Nature 2009;457(7225):102-6.

32. Balendiran GK, Dabur R, Fraser D. The role of glutathione in cancer. Cell Biochem Funct 2004;22(6):343-52.

33. Oyama T, Kagawa N, Kunugita N, Kitagawa K, Ogawa M, Yamaguchi T, et al. Expression of cytochrome P450 in tumor tissues and its association with cancer development. Front Biosci 2004;9:1967-76.

34. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. Oncogene 2006;25(11):1679-91.

35. Hase K, Shatney C, Johnson D, Trollope M, Vierra M. Prognostic value of tumor "budding" in patients with colorectal cancer. Dis Colon Rectum 1993;36(7):627-35.

36. Ueno H, Murphy J, Jass JR, Mochizuki H, Talbot IC. Tumour 'budding' as an index to estimate the potential of aggressiveness in rectal cancer. Histopathology 2002;40(2):127-32.

37. Mitrovic B, Schaeffer DF, Riddell RH, Kirsch R. Tumor budding in colorectal carcinoma: time to take notice. Mod Pathol 2012;25(10):1315-25.

38. Takahashi Y, Ishii G, Taira T, Fujii S, Yanagi S, Hishida T, et al. Fibrous stroma is associated with poorer prognosis in lung squamous cell carcinoma patients. J Thorac Oncol 2011;6(9):1460-7.

39. Masuda R, Kijima H, Imamura N, Aruga N, Nakamura Y, Masuda D, et al. Tumor budding is a significant indicator of a poor prognosis in lung squamous cell carcinoma patients. Mol Med Rep 2012;6(5):937-43.

40. Bronsert P, Enderle-Ammour K, Bader M, Timme S, Kuehs M, Csanadi A, et al. Cancer cell invasion and EMT marker expression: a three-dimensional study of the human cancer-host interface. J Pathol 2014;234(3):410-22.

41. Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016;7:12474.

Tables

Table 1: Patient cohorts. Summary of clinico-pathologic parameters including OS and RFS for TMA, WS and TCGA datasets. CI=95% confidence interval; TF computer/human=tumor fragmentation scores respectively computed or assessed by eye.

Patient cohorts	TMA		WS		TCGA	
Age						
median (range)	66	(18 - 87)	65	(40 - 87)	68	(39-90)
Sex						
female	42	(20%)	19	(19%)	83	(25%)
male	166	(80%)	80	(81%)	252	(75%)
pT						
1	43	(21%)	24	(24%)	68	(20%)
2	107	(51%)	29	(29%)	213	(64%)
3	43	(21%)	34	(35%)	40	(12%)
4	15	(7%)	12	(12%)	14	(4%)
pN						
0	110	(53%)	47	(48%)	209	(63%)
1	73	(35%)	31	(31%)	88	(27%)
2 - 3	25	(12%)	21	(21%)	35	(10%)
pM						
0	196	(94%)	95	(96%)	284	(98%)
1a/b	12	(6%)	4	(4%)	5	(2%)
Stage						
I	71	(34%)	24	(24%)	165	(49%)
II	76	(37%)	35	(35%)	108	(32%)
III	50	(24%)	36	(37%)	57	(17%)
IV	11	(5%)	4	(4%)	5	(2%)
Grade						
1 - 2	121	(58%)	44	(44%)	-	-
3	87	(42%)	55	(56%)	-	-
Size (cm)						
median (range)	3.9	(0.4 - 14.0)	5	(1.1 - 11.0)	-	-
Vessel infiltration						
absent	122	(59%)	52	(53%)	-	-
present	86	(41%)	47	(47%)	-	-
Adjuvant therapy						
no	133	(64%)	69	(70%)	304	(91%)
yes	75	(36%)	30	(30%)	31	(9%)
TF score						
Computer (IHC)						
median (range)	7	(2-38)	984	(163-6093)		
Human (H&E)						
median (range)	6	(2-40)	25	(2-132)	12	(1-68)
OS (months)						
median (CI)	42	(29 - 55)	52	(36 - 69)	56	(43-69)
follow-up max	169		137		157	
RFS (months)						
median (CI)	59	(40-78)	46	(27-65)	75	(45-105)
follow-up max	169		137		154	
Total patients (n=642)	**208**		**99**		**335**	

Table 2: Association of TF scores on whole sections with overall survival. Survival analysis by Cox univariate and multivariate regressions. HR=hazard ratio; CI=95% confidence interval. Computer: automatic scoring on pan-cytokeratin IHC stained tissue; Human: fragmentation scored by eye; TF (low/high): scores dichotomized at the median (categorical); TF (continuous): score as continuous variable.

OS		univariate			multivariate	
TMA (n=208)	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
TF (low/high) computer IHC	1.56	(1.12-2.16)	**0.008**	-	-	-
TF(continuous) computer IHC	1.05	(1.02-1.07)	**<0.001**	1.03	(1.01-1.06)	**0.018**
Stage [I, II, III, IV]	1.56	(1.29-1.90)	**<0.001**	1.45	(1.19-1.76)	**<0.001**
Grade [1,2,3]	1.34	(0.97-1.85)	0.075	1.20	(0.87-1.66)	0.266
Vessel infiltration	2.63	(1.89-3.65)	**<0.001**	1.64	(1.64-3.25)	**<0.001**
WS (n=99)	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
TF (low/high) computer IHC	2.06	(1.26-3.37)	**0.004**	-	-	-
TF(continuous) computer IHC	1.00	(1.00-1.00)	**<0.001**	1.00	(1.00-1.00)	**0.020**
Stage [I, II, III, IV]	1.81	(1.31-2.49)	**<0.001**	1.54	(1.08-2.20)	**0.016**
Grade [1,2,3]	1.76	(1.10-2.82)	**0.018**	1.45	(0.88-2.39)	0.144
Vessel infiltration	1.99	(1.23-3.24)	**0.005**	1.22	(0.71-2.09)	0.480
TCGA (n=335)	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
TF (low/high) human H&E	1.97	(1.40-2.77)	**<0.001**			
TF (continuous) human H&E	1.02	(1.01-1.03)	**0.001**	1.02	(1.01-1.03)	**0.001**
Stage [I, II, III, IV]	1.25	(1.02-1.53)	**0.030**	1.26	(1.03-1.54)	**0.025**
Grade [1,2,3]	-	-	-	-	-	-
Vessel infiltration	-	-	-	-	-	-

Figure Legends

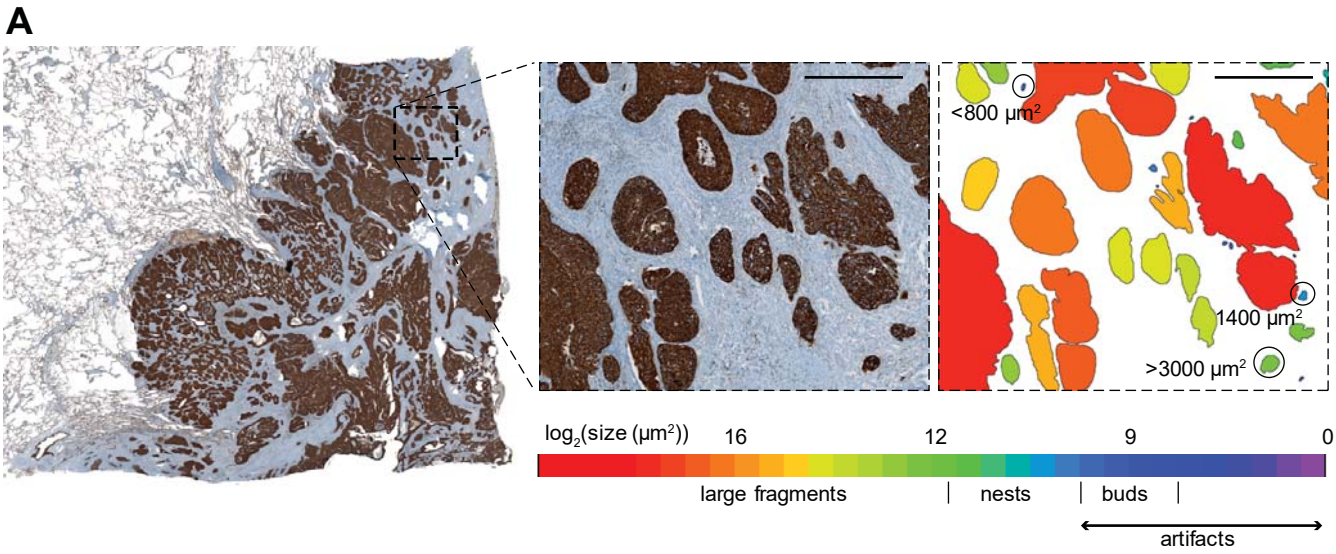
Figure 1. Computer-based morphometric analysis of tumor fragmentation. A) Left: histological whole section showing a squamous cell carcinoma of the lung stained with pan-cytokeratin IHC (brown) and its surrounding stroma (blue-grey counterstain). Right: color-based segmentation showing tumor fragments of different sizes labelled on a 32-colors scale. Approximate sizes are: single cell 200µm^2, buds ≤800µm^2, nests ≤3000µm^2, larger fragments >3000µm^2. Number of epithelial elements=32, size$_{mean}$=44'000µm^2, size$_{median}$=14'000µm^2. Scale bar: 500µm. B). Left: X-ray microtomography 2D reconstructed slice of a lung SCC example. Arrows indicate epithelial tumor fragments, t*=tumor, n*=necrosis, s*=stroma. Right: 3D rendering of the tumor microarchitecture imaged by X-ray microtomography.

Figure 2. Lung squamous cell carcinoma microarchitecture and fragmentation. Histological examples of pan-cytokeratin stained sections from TMA (A-B) and WS (C-D) cohorts. E-F) H&E stained sections from TCGA. Left/right: low/high fragmentation.

Figure 3. Overall survival curves based on tumor fragmentation. Kaplan Meier curves stratified by high/low TF, dichotomized at the median. HR=hazard ratio; CI=95% confidence interval. Computed TF scores were quantified on pan-cytokeratin stained tissue; Human-based scores were performed on H&E stained tissue.

Figure 4. Versican and Periostin immunohistochemistry. Boxplots showing IHC scores for Versican and Periostin compartmentalized expression in tumor and stroma grouped by high and low TF scores dichotomized at the median. B) Matching IHC stained tumor cores. Scale bar: 100µm.

Author Manuscript Published OnlineFirst on March 31, 2017; DOI: 10.1158/0008-5472.CAN-16-2363
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.



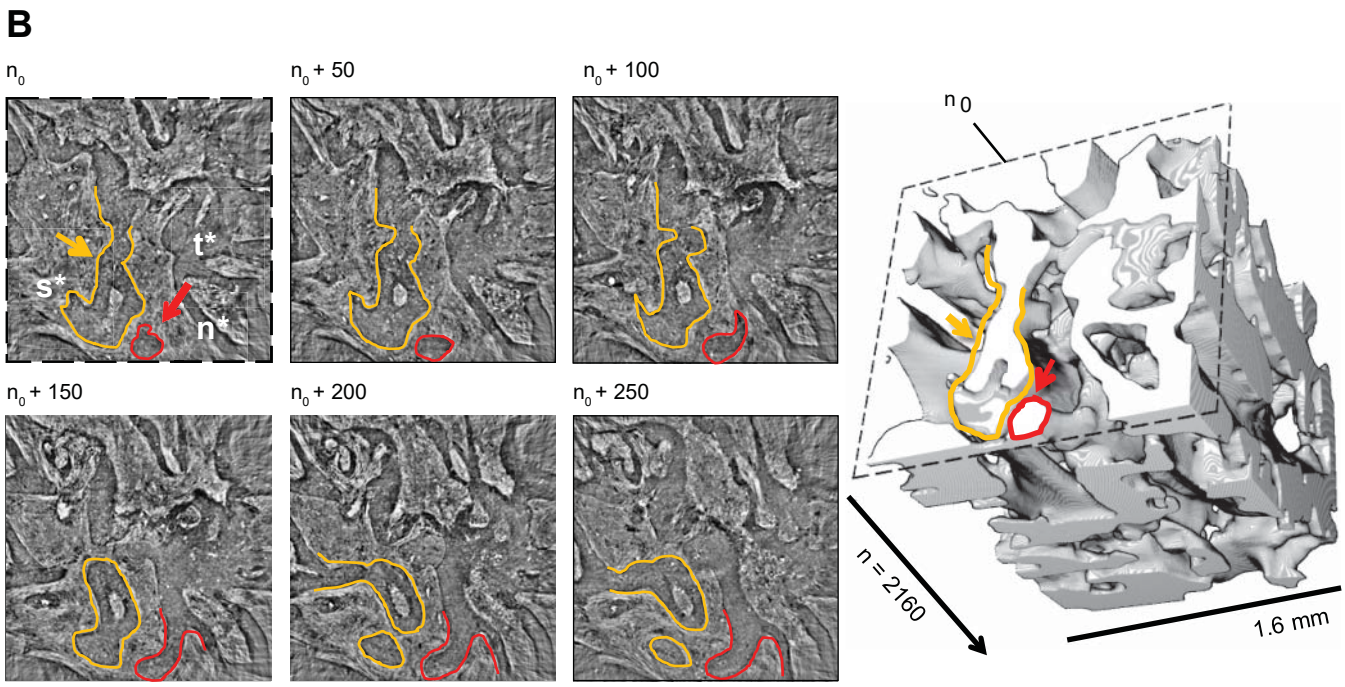


Figure 1

Author Manuscript Published OnlineFirst on March 31, 2017; DOI: 10.1158/0008-5472.CAN-16-2363
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

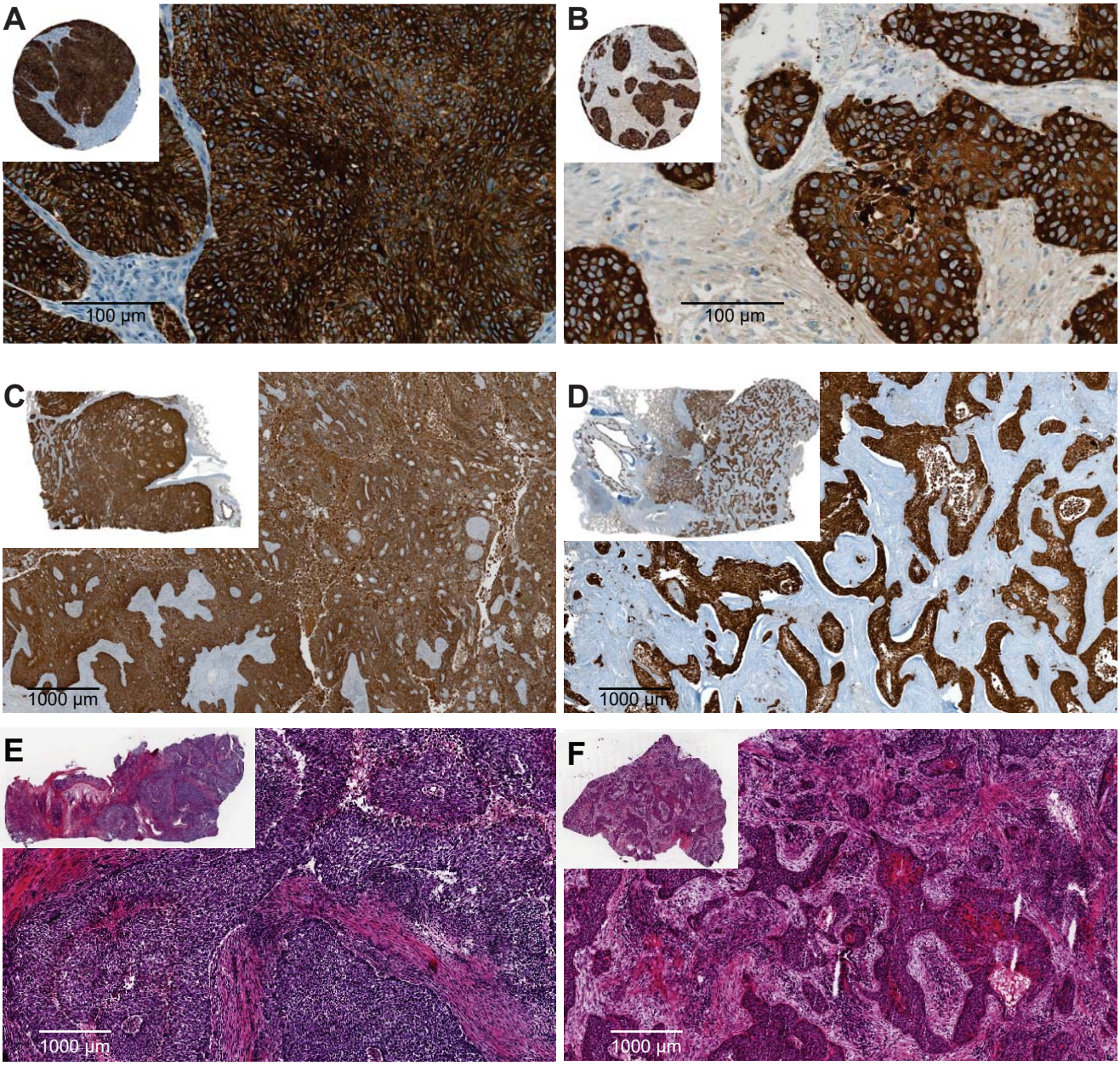


Figure 2

Author Manuscript Published OnlineFirst on March 31, 2017; DOI: 10.1158/0008-5472.CAN-16-2363
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

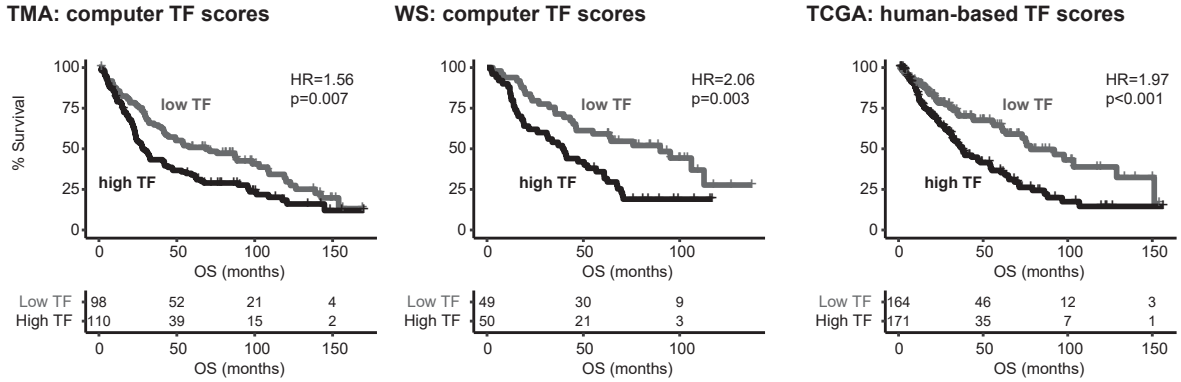


Figure 3

Author Manuscript Published OnlineFirst on March 31, 2017; DOI: 10.1158/0008-5472.CAN-16-2363
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

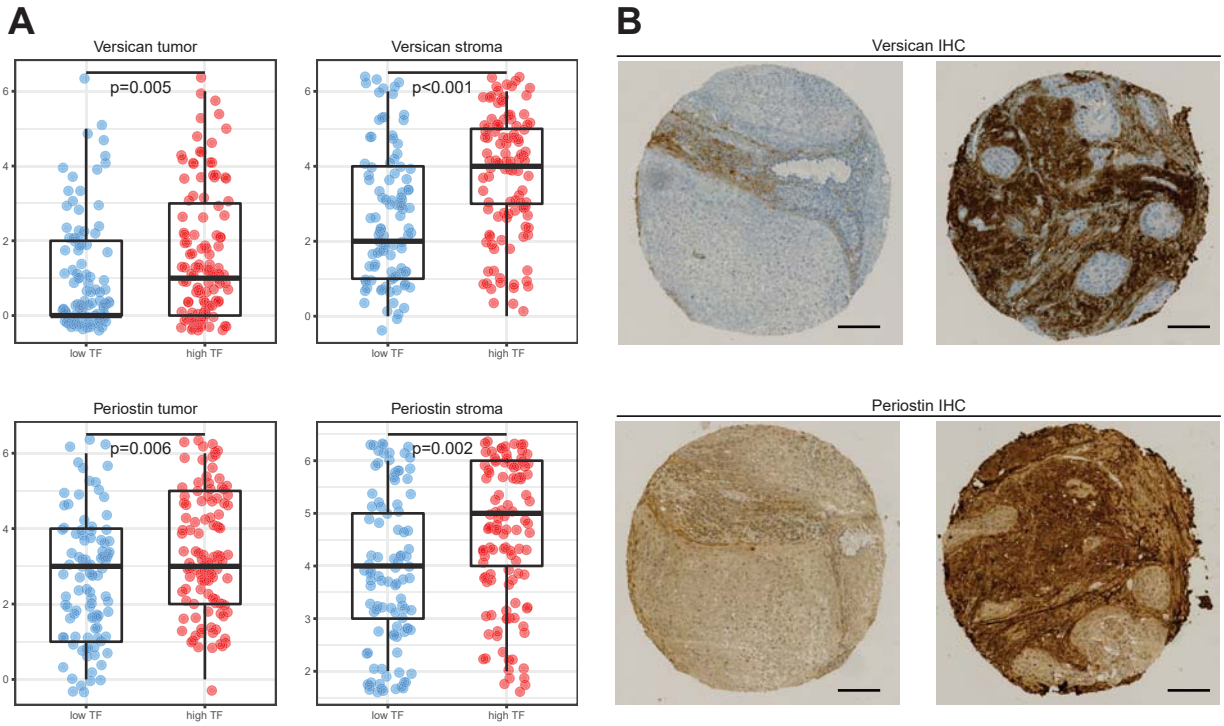


Figure 4

AACR American Association for Cancer Research

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Morphoproteomic characterization of lung squamous cell carcinoma fragmentation, a histological marker of increased tumor invasiveness

Ruben Casanova, Daniel Xia, Undine Rulle, et al.

Cancer Res Published OnlineFirst March 31, 2017.